Absence of epidemic threshold in scale-free networks with connectivity correlations

Marián Boguñá,[1] Romualdo Pastor-Satorras,[2] and Alessandro Vespignani[3]

[1]*Departament de Física Fonamental, Universitat de Barcelona, Av. Diagonal 647, 08028 Barcelona, Spain*
[2]*Departament de Física i Enginyeria Nuclear, Universitat Politècnica de Catalunya, Campus Nord, 08034 Barcelona, Spain*
[3]*The Abdus Salam International Centre for Theoretical Physics, P.O. Box 586, Trieste, I-34014, Italy*

(Dated: February 6, 2008)

Random scale-free networks have the peculiar property of being prone to the spreading of infections. Here we provide an exact result showing that a scale-free connectivity distribution with diverging second moment is a sufficient condition to have null epidemic threshold in unstructured networks with either assortative or dissortative mixing. Connectivity correlations result therefore ininfluential for the epidemic spreading picture in these scale-free networks. The present result is related to the divergence of the average nearest neighbors connectivity, enforced by the connectivity detailed balance condition.

PACS numbers: 89.75.-k, 87.23.Ge, 05.70.Ln

Complex networks play a capital role in the modeling of many social, natural, and technological systems which are characterized by peculiar topological properties [1, 2]. In particular, small-world properties [3] and scale-free connectivity distributions [4] appear as common features of many real-world networks. The statistical physics approach has been proved a very valuable tool for the study of these networks, and several surprising results concerning dynamical processes taking place on complex networks have been repeatedly reported. In particular, the absence of the percolation [5, 6] and epidemic [7, 8, 9, 10] thresholds in scale-free (SF) networks has hit the community because of its potential practical implications. The absence of the percolative threshold, indeed, prompts to an exceptional tolerance to random damages [11]. On the other hand, the lack of any epidemic threshold makes SF networks the ideal media for the propagation of infections, bugs, or unsolicited informations [7].

Recent studies have focused in a more detailed topological characterization of several social and technological networks. In particular, it has been recognized that many of these networks possess, along with SF properties, nontrivial connectivity correlations [12]. For instance, many social networks show that nodes with high connectivity will connect more preferably to highly connected nodes [12]; a property referred to as "assortative mixing". On the opposite side, many technological and biological networks show "dissortative mixing"; *i.e.* highly connected nodes are preferably connected to nodes with low connectivity [13, 14, 15]. Correlations are very important in determining the physical properties of these networks [16] and several recent works are addressing the effect of "dissortative mixing" correlations in epidemic spreading [17, 18, 19]. The fact that highly connected nodes (hubs) are more likely to transmit the infection to poorly connected nodes could somehow slow down the spreading process. By numerical simulations and analytical arguments it has been claimed that, if strong enough, connectivity correlations might reintroduce an epidemic threshold in SF networks, thus restoring the standard tolerance to infections.

In this paper we analyze in detail the conditions for the lack of an epidemic threshold in the susceptible-infected-susceptible model [20] in SF networks. We find the exact result that *a SF connectivity distribution $P(k) \sim k^{-\gamma}$ with $2 < \gamma \leq 3$ in unstructured networks with assortative or dissortative mixing is a sufficient condition for a null epidemic threshold in the thermodynamic limit.* In other words, the presence of two-point connectivity correlations does not alter the extreme weakness of SF networks to epidemic diffusion. This result is related to the divergence of the nearest neighbors average connectivity, divergence that is ensured by the connectivity detailed balance condition, to be satisfied in physical networks. The present analysis can be easily generalized to more sophisticated epidemic models.

In the following we shall consider unstructured undirected SF networks, in which the only relevant property of the nodes is their connectivity [21], with distribution $P(k) \sim Ck^{-\gamma}$, with $2 < \gamma \leq 3$. $P(k)$ is defined as the probability that a randomly selected node has k connections to other nodes. In this case the network has unbounded connectivity fluctuations, signalled by a diverging second moment $\langle k^2 \rangle \to \infty$ in the thermodynamic limit $k_c \to \infty$, where k_c is the maximum connectivity of the network. It is worth recalling that in growing networks k_c is related to the network size N as $k_c \sim N^{1/(\gamma-1)}$ [2]. Finally, we shall consider that the network presents assortative or dissortative mixing allowing for non trivial two-point connectivity correlations. This corresponds to allow a general form for the conditional probability, $P(k' \,|\, k)$, that a link emanated by a node of connectivity k points to a node of connectivity k'.

As a prototypical example for examining the properties of epidemic dynamics in SF networks we consider the susceptible-infected-susceptible (SIS) model [20], in which each node represents an individual of the popu-

lation and the links represent the physical interactions among which the infection propagates. Each individual can be either in a susceptible or infected state. Susceptible individuals become infected with probability λ if at least one of the neighbors is infected. Infected nodes, on the other hand, recover and become susceptible again with probability one. A different recovery probability can be considered by a proper rescaling of λ and the time. This model is conceived for representing endemic infections which do not confer permanent immunity, allowing individuals to go through the stochastic cycle susceptible $\rightarrow$ infected $\rightarrow$ susceptible by contracting the infection over and over again. In regular homogeneous networks, in which each node has more or less the same number of connections, $k \simeq \langle k \rangle$, it is possible to understand the behavior of the model by looking at the average density of infected individuals $\rho(t)$ (the prevalence). It is found that for a spreading probability $\lambda \geq \lambda_c$, where λ_c is the epidemic threshold depending on the network average connectivity and topology, the system reaches an endemic state with a finite stationary density ρ. If $\lambda \leq \lambda_c$, the system falls in a finite time in a healthy state with no infected individuals ($\rho = 0$).

In SF networks the average connectivity is highly fluctuating and the approximation $k \simeq \langle k \rangle$ is totally inadequate. To take into account the effect of the connectivity fluctuations, it has been shown that it is appropriate to consider the quantity ρ_k [7, 8, 16], defined as the density of infected nodes within each connectivity class k. This description assumes that the network is unstructured and that the classification of nodes according only to their connectivity is meaningful [21]. Following Ref. [16], the mean-field rate equations describing the system can be written as

$$\frac{d\rho_k(t)}{dt} = -\rho_k(t) + \lambda k \left[1 - \rho_k(t)\right] \sum_{k'} P(k' \,|\, k) \rho_{k'}(t). \quad (1)$$

The first term on the r.h.s. represents the annihilation of infected individuals due to recovery with unitary rate. The creation term is proportional to the density of susceptible individuals, $1 - \rho_k$, times the spreading rate, λ, the number of neighboring nodes, k, and the probability that any neighboring node is infected. The latter is the average over all connectivities of the probability $P(k' \,|\, k)\rho_{k'}$ that a link emanated from a node with connectivity k points to an infected node with connectivity k'. It is worth remarking that, while keeping into account the two point connectivity correlations, as given by the conditional probability $P(k' \,|\, k)$, yet we have neglected higher order density-density and connectivity correlations. Eq. (1) is therefore exact for the class of Markovian networks [16], in the limit of low prevalence ($\rho(t) \ll 1$).

In the case of uncorrelated networks each link points, with probability proportional to $k'P(k')$, to a node of connectivity k', regardless of the emanating node's connectivity. In this case, in the stationary state ($\partial_t \rho = 0$), $\sum_{k'} P(k' \,|\, k)\rho_{k'}(t)$ assumes a constant value independent on k and t and the system (1) can be solved self-consistently obtaining that the epidemic threshold is given by [10]

$$\lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle}. \quad (2)$$

For infinite SF networks with $\gamma \leq 3$, we have $\langle k^2 \rangle = \infty$, and correspondingly $\lambda_c = 0$; *i.e.* uncorrelated SF networks allow a finite prevalence whatever the spreading rate λ of the infection. Finally, from the solution of ρ_k, one can compute the total prevalence ρ using the relation $\rho = \sum_k P(k)\rho_k$.

In the case of correlated networks the explicit solution of Eq. (1) is not generally accessible. However, it has been shown that the epidemic threshold is given by [16]

$$\lambda_c = \frac{1}{\Lambda_m}, \quad (3)$$

where Λ_m is the largest eigenvalue of the *connectivity matrix* $\mathbf{C}$, defined by $C_{kk'} = kP(k' \,|\, k)$. In Ref. [16] it has been shown how this general formalism recovers previous results for uncorrelated networks, obtaining that, in this case, $\Lambda_m = \langle k^2 \rangle / \langle k \rangle$. More generally, by looking at Eq. (3), the absence of an epidemic threshold corresponds to a divergence of the largest eigenvalue of the connectivity matrix $\mathbf{C}$ in the limit of an infinite network size $N \rightarrow \infty$. In order to provide some general statement on the conditions for such a divergence we can make use of the Frobenius theorem for non-negative irreducible matrices [22]. This theorem states the existence of the largest eigenvalue of any non-negative irreducible matrix, eigenvalue which is simple, positive, and has a positive eigenvector. One of the consequences of the theorem is that it provides a bound to such largest eigenvalue [23]. In our case the matrix of interest is the connectivity matrix and, since $\mathbf{C}$ is non-negative and irreducible [24], it is possible to find lower and upper bounds of Λ_m. In particular, we can write [23]

$$\Lambda_m^2 \geq \min_k \sum_{k'} \sum_{\ell} k'\ell P(\ell \,|\, k) P(k' \,|\, \ell). \quad (4)$$

This inequality relates the lower bound of the largest eigenvalue Λ_m to the connectivity correlation function and, as we shall see, allows to find a sufficient condition for the absence of the epidemic threshold.

In order to provide an explicit bound to the largest eigenvalue we must exploit the properties of the conditional probability $P(k' \,|\, k)$. A key relation holding for all physical networks is that all links must point from one node to another. This is translated in the connectivity detailed balance condition [16]

$$kP(k' \,|\, k)P(k) = k'P(k \,|\, k')P(k'), \quad (5)$$

which states that the total number of links pointing from nodes with connectivity k to nodes of connectivity k' must be equal to the total number of links that point from nodes with connectivity k' to nodes of connectivity k. This relation is extremely important since it constraints the possible form of the conditional probability $P(k' \,|\, k)$ once $P(k)$ is given. By multiplying by a k factor both terms of Eq. (5) and summing over k' and k, we obtain

$$\langle k^2 \rangle = \sum_{k'} k' P(k') \sum_k k P(k \,|\, k'), \tag{6}$$

where we have used the normalization conditions $\sum_k P(k) = \sum_{k'} P(k' \,|\, k) = 1$. The term $\overline{k}_{nn}(k', k_c) = \sum_k k P(k \,|\, k')$ defines the average nearest neighbor connectivity (ANNC) of nodes of connectivity k'. This is a quantity customarily measured in SF and complex networks in order to quantify degree-degree correlations [13, 14, 15]. The dependence on k_c is originated by the upper cut-off of the k-sum and it must be taken into account since it is a possible source of divergences in the thermodynamic limit. In SF networks with $2 < \gamma < 3$ we have that the second moment of the connectivity distribution diverges as $\langle k^2 \rangle \sim k_c^{3-\gamma}$ [25]. We thus obtain that

$$\sum_{k'} k' P(k') \overline{k}_{nn}(k', k_c) \simeq \frac{C}{(3-\gamma)} k_c^{3-\gamma}. \tag{7}$$

In the case of dissortative mixing [12], the function $\overline{k}_{nn}(k', k_c)$ is decreasing with k' and, since $k' P(k')$ is an integrable function, the l.h.s. of Eq. (7) has no divergence related to the sum over k'. This implies that the divergence must be contained in the k_c dependence of $\overline{k}_{nn}(k', k_c)$. In other words, the function $\overline{k}_{nn}(k', k_c) \to \infty$ for $k_c \to \infty$ in a non-zero measure set. In the case of assortative mixing, $\overline{k}_{nn}(k', k_c)$ is an increasing function of k' and, depending on its rate of growth, there may be singularities associated to the sum over k'. Therefore, this case has to be analyzed in detail. Let us assume that the ANNC grows as $\overline{k}_{nn}(k', k_c) \simeq \alpha k'^{\beta}$, $\beta > 0$, when $k' \to \infty$. If $\beta < \gamma - 2$, again there is no singularity related to the sum over k' and the previous argument for dissortative mixing holds. When $\gamma - 2 \leq \beta < 1$ there is a singularity coming from the sum over k' of the type $\alpha k_c^{\beta-(\gamma-2)}$. However, since Eq. (7) comes from an identity, the singularity on the l.h.s. must match both the exponent of k_c and the prefactor on the r.h.s. In the case $\gamma - 2 \leq \beta < 1$, the singularity coming from the sum is not strong enough to match the r.h.s. of Eq. (7) since $\beta - (\gamma - 2) < 3 - \gamma$. Thus, the function $\overline{k}_{nn}(k', k_c)$ must also diverge when $k_c \to \infty$ in a non-zero measure set. Finally, when $\beta > 1$ the singularity associated to the sum is too strong, forcing the prefactor to scale as $\alpha \simeq r k_c^{1-\beta}$ and the ANNC as $\overline{k}_{nn}(k', k_c) \simeq r k_c^{1-\beta} k'^{\beta}$. It is easy to realize that $r \leq 1$, since the ANNC cannot be larger than k_c. Plugging the $\overline{k}_{nn}(k', k_c)$ dependence into Eq. (7) and simplifying common factors, we obtain the identity at the level of prefactors

$$\frac{r}{2 - \gamma + \beta} = \frac{1}{3 - \gamma}. \tag{8}$$

Since $\beta > 1$ and $r < 1$, the prefactor in the l.h.s. of Eq. (8) is smaller than the one of the r.h.s. This fact implies that the tail of the distribution in the l.h.s. of Eq. (7) cannot account for the whole divergence of its r.h.s. This means that the sum is not the only source of divergences and, therefore, the ANNC must diverge at some other point [26].

The large k_c behavior of the ANNC can be plugged in Eq. (4) obtaining that

$$\Lambda_m^2 \geq \min_k \sum_{\ell} \ell P(\ell \,|\, k) \overline{k}_{nn}(\ell, k_c) \tag{9}$$

The r.h.s. of this equation is a sum of positive terms and diverges with k_c at least as $\overline{k}_{nn}(\ell, k_c)$ both in the dissortative or assortative cases [27]. This readily implies that $\Lambda_m \geq \infty$ for all networks with diverging $\langle k^2 \rangle$. Finally Eq. (3) yields that the epidemic threshold vanishes in the thermodynamic limit in all SF networks with assortative and dissortative mixing if the connectivity distribution has a diverging second moment; *i.e. a SF connectivity distribution with exponent* $2 < \gamma \leq 3$ *is a sufficient condition for the absence of an epidemic threshold in unstructured networks with arbitrary two-point connectivity correlation function.*

In physical terms, the absence of the epidemic threshold is related to the divergence of the average nearest neighbors connectivity $\langle \overline{k}_{nn} \rangle_N$ in SF networks. This function is defined by

$$\langle \overline{k}_{nn} \rangle_N = \sum_k P(k) \overline{k}_{nn}(k, k_c), \tag{10}$$

where we have explicitly considered k_c as a growing function of the network size N. By using the analysis shown previously it follows that $\langle \overline{k}_{nn} \rangle_N \to \infty$ when $N \to \infty$. In SF networks this parameter takes into account the level of connectivity fluctuations and appears as ruling the epidemic spreading dynamics. Somehow the number of neighbors that can be infected in successive steps is the relevant quantity. Only in homogeneous networks, where $\langle \overline{k}_{nn} \rangle_N \simeq \langle k \rangle$, the epidemic spreading properties can be related to the average connectivity. Noticeably, the power-law behavior of SF networks imposes a divergence of $\langle \overline{k}_{nn} \rangle_N$ independently of the level of correlations present in the network. This amounts to lower to zero the epidemic threshold. On the practical side, connectivity correlation functions can be measured in several networks and show assortative or dissortative behavior depending on the system. These measurements are always performed in the presence of a finite k_c that allows

the regularization of the function $\overline{k}_{nn}(k, k_c)$. The most convenient way to exploit the infinite size singularity is to measure the average nearest neighbor connectivity for increasing network sizes. All SF networks with $2 < \gamma \leq 3$ must present a diverging $\left\langle \overline{k}_{nn} \right\rangle_N$ for $N \to \infty$. This statement is independent of the structure of the correlations present in the networks.

It is worth stressing that the divergence of $\left\langle \overline{k}_{nn} \right\rangle_N$ is ensured by the connectivity detailed balance condition alone. Thus it is a very general results holding for all SF networks with $2 < \gamma \leq 3$. On the contrary, the SF behavior with $2 < \gamma \leq 3$ is a necessary condition for the lack of epidemic threshold only in networks with general two-point connectivity correlations and in absence of higher-order correlations. The reason is that the relation between the epidemic threshold and the maximum eigenvalue of the connectivity matrix only holds for these classes of networks. Higher order correlations, or the presence of an underlying metric in the network [21], can modify the rate equation at the basis of the SIS model and may invalidate the present discussion.

This work has been partially supported by the European commission FET Open project COSIN IST-2001-33555. R.P.-S. acknowledges financial support from the Ministerio de Ciencia y Tecnología (Spain).

[1] R. Albert and A.-L. Barabási, Rev. Mod. Phys. **74**, 47 (2002).

[2] S. N. Dorogovtsev and J. F. F. Mendes, Adv. Phys. **51**, 1079 (2002).

[3] D. J. Watts and S. H. Strogatz, Nature **393**, 440 (1998).

[4] A.-L. Barabási and R. Albert, Science **286**, 509 (1999).

[5] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. Lett. **85**, 5468 (2000).

[6] R. Cohen, K. Erez, D. ben Avraham, and S. Havlin, Phys. Rev. Lett. **86**, 3682 (2001).

[7] R. Pastor-Satorras and A. Vespignani, Phys. Rev. Lett. **86**, 3200 (2001).

[8] R. Pastor-Satorras and A. Vespignani, Phys. Rev. E **63**, 066117 (2001).

[9] R. M. May and A. L. Lloyd, Phys. Rev. E **64**, 066112 (2001).

[10] R. Pastor-Satorras and A. Vespignani, in *Handbook of Graphs and Networks: From the Genome to the Internet*, edited by S. Bornholdt and H. G. Schuster (Wiley-VCH, Berlin, 2002), e-print cond-mat/0205260.

[11] R. A. Albert, H. Jeong, and A.-L. Barabási, Nature **406**, 378 (2000).

[12] M. E. J. Newman, *Assortative mixing in networks* (2002), e-print cond-mat/0205405.

[13] R. Pastor-Satorras, A. Vázquez, and A. Vespignani, Phys. Rev. Lett. **87**, 258701 (2001).

[14] A. Vázquez, R. Pastor-Satorras, and A. Vespignani, Phys. Rev. E **65**, 066130 (2002).

[15] S. Maslov and K. Sneppen, Science **296**, 910 (2002).

[16] M. Boguñá and R. Pastor-Satorras, *Epidemic spreading in correlated complex networks* (2002), e-print cond-mat/0205621.

[17] V. M. Eguíluz and K. Klemm, *Epidemic threshold in structured scale-free networks* (2002), e-print cond-mat/0205439.

[18] C. P. Warren, L. M. Sander, and I. M. Sokolov, *Geography in a scale-free network model* (2002), e-print cond-mat/0207324.

[19] P. Blanchard, C.-H. Chang, and T. Krueger, *Epidemic thresholds on scale-free graphs: the interplay between exponent and preferential choice* (2002), e-print cond-mat/0207319.

[20] R. M. Anderson and R. M. May, *Infectious diseases in humans* (Oxford University Press, Oxford, 1992).

[21] Our argument does not apply for structured networks in which a distance can be defined and therefore do not show the small-world property, *e. g.* K. Klemm and V. M. Eguíluz, Phys. Rev. E **65**, 036123 (2002).

[22] F. R. Gantmacher, *The theory of matrices*, vol. II (Chelsea Publishing Company, New York, 1974).

[23] From the Frobenius theorem [22] it can be proved that the maximum eigenvalue, Λ_m, of any non-negative irreducible matrix, $A_{kk'}$, satisfies the inequality $\Lambda_m \geq \min_k \frac{1}{\psi(k)} \sum_{k'} A_{kk'} \psi(k')$, where $\psi(k)$ is any positive vector. By setting $\mathbf{A} = \mathbf{C}^2$ and $\psi(k) = k$ we recover the inequality of Eq. (4).

[24] The irreducible property of the connectivity matrix is a simple consequence of the fact that all the connectivity classes in the network are accessible. That is, starting from the connectivity class k it is always possible to find a path of links that connects this class to any other class k' of the network. If this is not the case it means that the network is built up of disconnected irreducible subnetworks and, therefore, we can apply the same line of reasoning to each subnetwork. Notice that being irreducible is not equivalent to being fully connected at the node to node level, but at the class to class level.

[25] For $\gamma = 3$ the second moment diverges as $\left\langle k^2 \right\rangle \sim \ln k_c$ but the argument, though more involved, is still valid.

[26] The case $\beta = 1$, that is, $\overline{k}_{nn}(k', k_c) \sim \alpha k'$ when $k' \to \infty$, is rather pathological because is the only case where the ANNC can be a non diverging function. If $\alpha < 1$, Eq. (7) implies that it must exist at least one diverging point out of the one coming from the sum over k' and the discussion made in the text is valid. However, when $\alpha = 1$ this divergence is enough to fulfill the identity (7) and $\overline{k}_{nn}(k, k_c)$ can be a convergent function when $k_c \to \infty$. Even in this pathological situation it is possible to prove that the maximum eigenvalue diverges in the thermodynamic limit. In order to do so the argumentation is to be translated to the second moment of the nearest neighbor's connectivity. Since the inequality $\langle x^2 \rangle \geq \langle x \rangle^2$ is true for any random variable, the function $\overline{k^2}_{nn}(k, k_c)$ is either $\sim k'^2$, in which case the maximum eigenvalue is trivially ∞, or the probability $P(k' \,|\, k)$ approaches to SF distribution when $k_c \to \infty$ and Λ_m diverges as well.

[27] One may argue that, since we are calculating a minimum for k, if the transition probability $P(\ell \,|\, k_0)$ is zero at some point k_0, this minimum is zero. In this case it is possible to show that repeating the same argument with $\mathbf{C}^3$ instead of $\mathbf{C}^2$ provides us an inequality that avoids this problem.